

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-097541
 (43)Date of publication of application : 14.04.1998

(51)Int.CL G06F 17/30
 G06F 3/14
 G06F 12/00
 G06F 13/00
 G06T 1/00

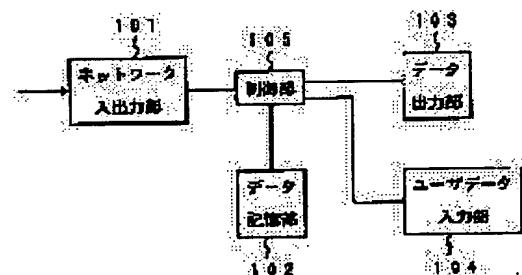
(21)Application number : 08-247799 (71)Applicant : TOSHIBA CORP
 (22)Date of filing : 19.09.1996 (72)Inventor : HIRAKAWA HIDEKI
 KIMURA KAZUHIRO

(54) DOCUMENT BROWSING SUPPORT SYSTEM AND DOCUMENT PROCESSING SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To give related information to the document of a browsing object by a person except for a document generator by generating and outputting output information corresponding to a detected expression pattern from the document on a computer network, based on related information in data.

SOLUTION: A resource on the computer network is accessed through a network input/output part 101 in accordance with the access request of effect that the specified document is to be referred to. Data received in the network input/output part 101 is transferred to a data output part 108 through a control part 105. Then, at least the expression pattern in the document and related information related to output information for obtaining output information corresponding to the pattern is described are inputted. The expression pattern in data is referred to and the expression pattern in the document on the computer network is detected. Output information corresponding to the expression pattern detected based on related information is generated from the document on the computer network and it is outputted.



LEGAL STATUS

[Date of request for examination] 29.08.2000
 [Date of sending the examiner's decision of rejection]
 [Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
 [Date of final disposal for application]
 [Patent number] 3195252

[Date of registration] 01.06.2001

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平10-97541

(43)公開日 平成10年(1998)4月14日

(51)Int.Cl.⁶
G 0 6 F 17/30
3/14 3 6 0
12/00 5 4 7
13/00 3 5 4
G 0 6 T 1/00

識別記号

F I
G 0 6 F 15/419
3/14
12/00
13/00
15/40

3 2 0
3 6 0 C
5 4 7 H
3 5 4 D
3 7 0 A

審査請求 未請求 請求項の数 4 O L (全 11 頁) 最終頁に統ぐ

(21)出願番号 特願平8-247799
(22)出願日 平成8年(1996)9月19日

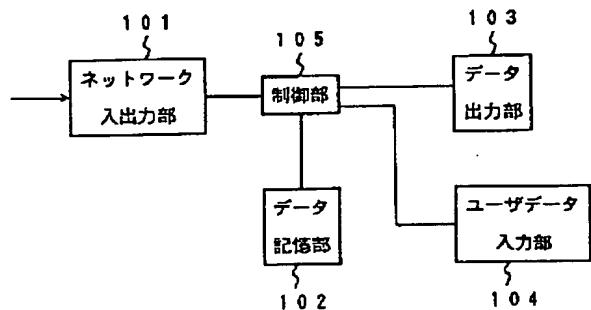
(71)出願人 000003078
株式会社東芝
神奈川県川崎市幸区堀川町72番地
(72)発明者 平川 秀樹
神奈川県川崎市幸区小向東芝町1番地 株式会社東芝研究開発センター内
(72)発明者 木村 和広
神奈川県川崎市幸区小向東芝町1番地 株式会社東芝研究開発センター内
(74)代理人 弁理士 鈴江 武彦 (外6名)

(54)【発明の名称】 文書ブラウズ支援システムおよび文書処理システム

(57)【要約】

【課題】文書作成者以外の人が作成した関連情報をコンピュータネットワーク上の閲覧対象の文書に付与できるような環境を実現する文書ブラウズ支援システムを提供する。

【解決手段】文書中の表現パタンとそれに対応する出力情報を得るための該出力情報に関連する関連情報を記述したデータを入力し、このデータ中の表現パタンを参照してコンピュータネットワーク上の文書中の表現パタンを検出し、入力されたデータ中の関連情報を基に、検出された表現パタンに対応する出力情報をコンピュータネットワーク上の文書から生成して出力する。



1

【特許請求の範囲】

【請求項1】コンピュータネットワーク上の文書を閲覧するための文書ブラウジングシステムに適用される文書ブラウズ支援システムにおいて、少なくとも文書中の表現パタンとそれに対応する出力情報を得るための該出力情報を関連する関連情報を記述したデータを入力する入力手段と、

前記入力手段により入力されたデータ中の表現パタンを参照して前記コンピュータネットワーク上の文書中の表現パタンを検出する検出手段と、

前記データ入力手段により入力されたデータ中の関連情報を基に、前記検出手段により検出された表現パタンに対応する出力情報を前記コンピュータネットワーク上の文書から生成して出力する出力手段とを備えたことを特徴とする文書ブラウズ支援システム。

【請求項2】コンピュータネットワーク上の文書を閲覧するための文書ブラウジングシステムに適用される文書ブラウズ支援システムにおいて、

少なくとも文書中の表現パタンとそれに対応する出力情報を得るための該出力情報を関連する関連情報を記述したデータを入力する入力手段と、

前記入力手段により入力されたデータ中の表現パタンを参照して前記コンピュータネットワーク上の文書中の表現パタンを検出する検出手段と、

前記データ入力手段により入力されたデータ中の関連情報を基に、前記検出手段により検出された表現パタンに対応する出力情報を前記コンピュータネットワーク上の文書から生成して出力する出力手段と、

前記データ入力手段により入力されたデータ中の関連情報を基に、前記コンピュータネットワーク上の前記検出手段により検出された表現パタンを含む文書の少なくとも一部を抽出して記憶する記憶手段とを備えたことを特徴とする文書ブラウズ支援システム。

【請求項3】コンピュータネットワーク上の文書を閲覧するための文書ブラウジングシステムに適用される文書ブラウズ支援システムにおいて、

少なくとも文書中の表現パタンとコンピュータネットワーク上の該表現パタンを含む文書の位置を示す文書位置情報をと対応付けて記述したデータを入力する入力手段と、

前記入力手段により入力されたデータ中の表現パタンを参照して前記コンピュータネットワーク上の文書中の表現パタンを検出する検出手段と、

前記検出手段により検出された表現パタンに対応する文書位置情報を前記入力手段により入力されたデータから検索する検索手段と、

前記検索手段により検索された文書位置情報を出力する出力手段とを備えたことを特徴とする文書ブラウズ支援システム。

【請求項4】文字列とコンピュータネットワーク上の該

50

2

文字列を含む文書の位置を示す文書位置情報をと対応付けて格納した情報格納手段と、

入力された文字列を基に前記情報格納手段から前記文書位置情報を検索する検索手段と、

前記検索手段により検索された文書位置情報を出力する出力手段とを備えたことを特徴とする文書処理システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、文書ブラウズ支援システムおよびこれに好適な文書処理システムに関する。

【0002】

【従来の技術】インターネットの普及に伴い、コンピュータネットワーク上の文書を閲覧できる文書ブラウジングシステム、いわゆる文書ブラウザが一般に利用されるようになっている。こうした文書ブラウザを利用するため、コンピュータネットワーク上のある文書に現れる語句などの情報に関する関連情報は、コンピュータネットワーク上の他の文書へのリンクがとられており、マウスでクリックするなどの簡単な操作で、他の文書を閲覧することができる。しかしながら、このようなリンクは予めブラウジングに供される文書作成者により用意されたものであり、必ずしもその文書の読者が望むような情報が用意されているわけではない。

【0003】

【発明が解決しようとする課題】このように、従来の文書ブラウズシステムにおいては、文書作成者が用意した語句などの情報のみについて関連情報が対応付けられており、文書作成者以外の人が作成した関連情報を閲覧対象の文書に付与することはできないという問題があった。

【0004】本発明は、文書作成者以外の人が作成した関連情報をコンピュータネットワーク上の閲覧対象の文書に付与できるような環境を実現する文書ブラウズ支援システムを提供することを目的とする。

【0005】

【課題を解決するための手段】上記の課題を解決するため、本発明はコンピュータネットワーク上の文書を閲覧するための文書ブラウジングシステムに適用される文書ブラウズ支援システムにおいて、少なくとも文書中の表現パタンとそれに対応する出力情報を得るための該出力情報を関連する関連情報を記述したデータを入力する入力手段と、この入力手段により入力されたデータ中の表現パタンを参照してコンピュータネットワーク上の文書中の表現パタンを検出する検出手段と、データ入力手段により入力されたデータ中の関連情報を基に、この検出手段により検出された表現パタンに対応する出力情報をコンピュータネットワーク上の文書から生成して出力する出力手段とを備えたことを特徴とする。

【0006】また、データ入力手段により入力されたデータ中の関連情報を基に、コンピュータネットワーク上の検出手段により検出された表現パタンを含む文書の少なくとも一部を抽出して記憶する記憶手段をさらに備えたことを特徴とする。

【0007】このような構成により、本発明によるとコンピュータネットワーク上の文書をブラウジングする際、文書作成者以外のユーザによりブラウジング対象の文書への関連情報の付与ができる。

【0008】さらに、本発明はコンピュータネットワーク上の文書を閲覧するための文書ブラウジングシステムに適用される文書ブラウズ支援システムにおいて、少なくとも文書中の表現パタンとコンピュータネットワーク上の該表現パタンを含む文書の位置を示す文書位置情報を対応付けて記述したデータを入力する入力手段と、この入力手段により入力されたデータ中の表現パタンを参照してコンピュータネットワーク上の文書中の表現パタンを検出する検出手段と、この検出手段により検出された表現パタンに対応する文書位置情報を入力手段により入力されたデータから検索する検索手段と、この検索手段により検索された文書位置情報を出力する出力手段とを備えたことを特徴とする。

【0009】このような構成によって、コンピュータネットワーク上のブラウジング対象の文書中に文書位置情報であるハイパーアリンクが存在しない場合でも、ハイパーアリンクを対応付けて出力することが可能となる。

【0010】本発明に係る文書処理システムは、文字列とコンピュータネットワーク上の該文字列を含む文書の位置を示す文書位置情報を対応付けて格納した情報格納手段と、入力された文字列を基に情報格納手段から文書位置情報を検索する検索手段と、この検索手段により検索された文書位置情報を出力する出力手段とを備えたことを特徴とする。

【0011】このようにして、ハイパーアリンクを含む文書を作成する場合、文書位置情報として他者が文書中に作成したハイパーアリンク情報を検索し、表示などの出力を行って利用することが可能となる。

【0012】

【発明の実施の形態】以下、図面を参照して本発明の実施形態を説明する。

(第1の実施形態) 図1は、本発明の一実施形態に係る文書ブラウズ支援システムの構成を示すブロック図である。この文書ブラウズ支援システムは、図示しないコンピュータネットワークよりデータを取り出したり、コンピュータネットワークへデータを送信する機能を持つネットワーク入出力部101と、各種データを記憶する機能を持つデータ記憶部102と、CRTディスプレイや液晶ディスプレイなどの表示手段を含みユーザへ各種情報を提示する機能を有するデータ出力部103と、キーボード、ペン入力装置、フロッピーディスクなどの記憶

媒体、あるいはコンピュータなど、ユーザが所望のデータを入力するためのユーザデータ入力部104と、これらのネットワーク入出力部101、データ記憶部10

2、データ出力部103およびユーザデータ入力部104を制御する制御部105とからなる。

【0013】まず、コンピュータネットワーク上に存在する文書を閲覧、すなわち任意の文書を取り出して表示する文書ブラウジング機能について簡単に説明する。これは現在広く使用されているWWWブラウザに見られる基本機能であり、以下の処理を行う。すなわち、ユーザから特定の文書を閲覧したい旨のアクセス要求を受けると、その要求に応じてネットワーク入出力部101を介してコンピュータネットワーク上のリソースにアクセスを行って、指定された文書のデータを要求する。この要求に対しコンピュータネットワークからデータが送信され、ネットワーク入出力部101で受信されると、そのデータを制御部105を介してデータ出力部103に転送し、ユーザに提示する。

【0014】本実施形態は、図2に概要を示すデータ(以下、これをプラグインパッケージという)をユーザデータ入力部104やネットワーク入出力部101などから入力して文書ブラウジングを行うことを特徴とするものである。このプラグインパッケージには、少なくとも文書中の表現パタンと、それに対応する出力情報を得るための出力情報に関連した関連情報を記述したデータである複数のパタンアクションユニット201、202、203が含まれている。ここで、出力情報とは例えば文書ブラウジング時にデータ出力部103で出力、例えれば表示したい情報である。

【0015】パタンアクションユニット201、202、203は、この例ではパタン記述部とアクション記述部およびステータス記述部の3つのパートから構成されている。これらの各パートの詳細については、後述する。また、プラグインパッケージのパッケージデータ記述部には、パッケージタイトル、日時、場所および個人情報などの情報が記述される。

【0016】本実施形態では、図3に示すアルゴリズムによってパタンアクションユニット201、202、203を参照実行する。以下、図2中の2つのパタンアクションユニット(パタンアクションユニット1、2)の具体例を示した図4を例にとり、図3のアルゴリズムについて説明する。図3のアルゴリズムは、ユーザからの指示によりコンピュータネットワークから取得された閲覧対象となる文書(以下、入力文書と呼ぶ)に対して適用される。

【0017】まず、ステップS01の前処理において、入力文書に対するマッチング処理のための前処理が行われる。この前処理は文書ブラウズ支援システムの適用例により種々異なるが、ここでは入力文書に対する辞書引き処理、形態素解析処理および文解析処理を行う。これ

5

ら処理の詳細は述べないが、機械翻訳システムなどの自然言語処理システムに適用されている公知の技術である。この前処理により、入力文書中の単語標準形（例えば、原文中の“going”に対して“go（動詞、ing形）”）、単語間の関係（例えば、原文“*I play tennis.*”に対して“*I*”が“play”的主体／主語、“*tennis*”が“play”的目的語／対象格などの構文／意味関係）が得られる。

【0018】次に、ステップS02のマッチングパターンチェックでは、図4のパターンアクションユニット1、2中のパターン記述部に記述された表現パターンと、入力文書中の表現パターンおよびこれをステップS01で前処理した表現パターンとのマッチングが行われる。すなわち、前処理した表現パターンを参照して入力文書中の表現パターンが検出される。

【0019】例えば、図4のパターンアクションユニット1のパターン記述部には、“*Why don't you —*”という表現パターンが記述されている。このパターンは、“*W:*”という識別子によって単語レベルでのマッチングを行うという旨の指示がされており、ステップS01の前処理で形態素解析された単語列に対するマッチング処理に供される。

【0020】この場合、ステップS01の前処理で入力文書中の単語の標準化処理がなされているため、入力文書中の表現が“*Why don't you dance with me ?*”のように大文字を含む単語（この例では*Why*）がある場合もマッチングが可能となる。また、語形変化、活用なども同様に標準化されているので、マッチングを行うことができる。マッチングの記述部には、ワイルドカードなど一般的なマッチング記述が可能である。

【0021】一方、図4のパターンアクションユニット2中のパターン記述部は、“*M:*”という識別子によってパターンが意味解析結果に供するものである旨が指示されているため、ステップS01の前処理で得られている、入力文書に対する意味表現（ここでは図示しない意味解析木）とのマッチングが行われる。このパターン記述部の“*ComputerCompany:X”は、意味カテゴリが「コンピュータメーカー」である語であり、その語を“*X*”として後述するプログラムで参照可能とするという効果を持つ。

意味カテゴリについては、既存の概念体系、シソーラス（図示しない）を利用することができる。また、このパターン記述部中の“*buy | acquire*”は、“*buy*”または“*acquire*”のいずれかの単語とマッチするという記述である。さらに、このパターン記述部中の“*agent*”，“*object*”は、単語間の意味関係を表す関係標識であり、“*HUL acquired Ratos*”，“*Ratos was acquired by HUL*”といった表現を含む文章の部分とマッチする。

【0022】このようにパターンアクションユニット中のパターン記述部には、単語の標準化や構文的・意味的抽象

6

化を行った、文書中の表現パターンが記述される。次に、ステップS03において、ステップS02のマッチングパターンチェックにより、パターンアクションユニット中のパターン記述部に記述された表現パターンとマッチする表現パターンが入力文書中に存在しているか否かをチェックし、存在している場合には、ステップS04の動作適用処理を行う。この動作適用処理では、パターンアクションユニット中のアクション記述部の記述に応じた処理を行う。

【0023】例えば、図4のパターンアクションユニット1中のアクション記述部には、「タグ設定（#1）」というアクションが記述されているが、これはパターン記述部に記述された表現パターンとマッチした入力文書中の表現パターンの部分に対して、#1というデータを利用したプログラム起動のためのリンク情報を文書に付与するということを示している。これは、WWW（World Wide Web）のHTML文書におけるリンクに相当するのと同様の効果を与えるものである。この結果、図4のパターンアクションユニット1の例では、入力文書中の“*Why don't you dance with me ?*”という文字列が、例えば“<Pref:Why don't you/タグ#1>dance with me ?”のように変換される。タグ#1の内容の一例は、図4の左下に記載した通りである。

【0024】ここで、“*Pref:*”タグが付加された<Pref: Why …>の部分は、データ出力部103において他の部分と区別できるように異なった様態、例えば異なった色で表示されることにより、この部分はパターン記述部に記述された表現パターンにマッチしていることをユーザが認識できる。そして、この<Pref: Why …>の部分をユーザが指示してクリックするなどの実行指示をユーザデータ入力部104により行うと、パターンアクションユニット1中のアクション記述部のデータに基づいて所定の処理が行われる。

【0025】一方、図4のパターンアクションユニット2では、アクション記述部に「プログラム#1」という記述がなされているが、これはユーザによる実行指示を待たずにアクション記述部の処理を行うことを示している。プログラム#1の内容の一例は、図4の右下に記載した通りである。

【0026】なお、ステップS04の動作適用処理では、パターンアクションユニット中のステータス記述部に関する処理も並行して行われる。ステータス記述部の第1フィールドは、ON/OFFの記述がなされる。これはパターンアクションユニット自体の動作適用処理を行うか、行わないかを制御するものであり、OFFの場合はパターン記述部に記述された表現パターンに入力文書中の表現パターンがマッチしても、動作適用処理は行われない。このON/OFFの指定は、ユーザが図1のユーザデータ入力部104により行う。

【0027】ステータス記述部の第2フィールドは、入

力文書中の表現パターンの出現頻度を記憶するものであり、パターン記述部に記述された表現パターンにマッチした表現パターンが入力文書中に出現する度にインクリメントする。この第2フィールドの内容を参照することにより、これがある所定の出現頻度を越えたパターンアクションユニットをONにする、すなわち出現頻度の高い表現パターンについてのみ、アクション記述部に記述された処理を行ったり、特定の出現頻度に達する毎に処理を行ったりすることが可能となる。

【0028】ステップS01～S04の処理は、ステップS05で入力文書の文書データの終りが見出されるまで繰り返し行われる。次に、具体的な文書の一例を用いて、本実施形態の文書ブラウズ支援システムの動作を述べる。

【0029】図5は、コンピュータネットワーク上の他の文書へのハイパーリンクを含むコンピュータネットワーク上の文書をネットワーク入出力部101を介して入力文書として取り込み、データ出力部103で表示した時の表示画面の一例である。下線は、ハイパーリンクが存在することを明示的に示している。この入力文書のソーステキストにおいては、ハイパーリンクの部分にそれを示すタグおよびコンピュータネットワーク上の文書へのポインタ（図示せず）が含まれている。このハイパーリンクを有する部分にユーザがマウスのカーソルを移動させ、クリックすることによって、コンピュータネットワーク上の対応する文書がネットワーク入出力部102を介してデータ記憶部102に入力文書として取り込まれ、制御部105によりデータ出力部103で表示される。

【0030】図5の“New Information Guide”的部分に対応する文書は、図6の画面（ラウズ画面）に示される文書であるとする。図5の文書は、上述のようにデータ記憶部102に既に記憶され、データ出力部103で表示されている。この表示画面上でユーザが“New Information Guide”的部分をクリックすると、そのクリックの情報がユーザデータ入力部104より制御部105へ渡される。制御部105は、これを受け対応する文書の所在情報をデータ記憶部102より取り出し、ネットワーク入出力部101を介してコンピュータネットワーク上の所定のコンピュータに対してその文書データの送出要求を出す。この送出要求を受けた要求先のコンピュータは、要求された文書データを要求元のコンピュータに送出する。

【0031】図7は、こうしてコンピュータネットワーク上の要求先のコンピュータから要求元のコンピュータに送出された文書データの一部分であり、図6に示されるラウズ画面の元となるデータである。この文書データは、HTML (Hyper Text Markup Language) のようなタグ付きのテキストである。この文書データが要求元のコンピュータに入力されると、まずタグの部分と表示内

容となるコンテンツの部分が解析され、コンテンツの部分が抽出される。この処理は、SGMLパーサ(Standard Generalized Markup Language parser)などの既存技術で行うことができる。

【0032】こうして抽出されたコンテンツの部分に対して、図3のステップS01の前処理が施される。この前処理により、前述したように“Why don't you try new products”という部分の単語標準形について、図8に示すような情報が得られる。第1番目の単語に関しては、“1”は単語の番号、“why”は入力単語“Why”的標準形、“(WH, HeadCap)”のWHは品詞であり、HeadCapは最初の文字が大文字化されていたという情報、45～47は原文中の単語Whyとの関連を保持するための情報である。ここでは図示しないが、ステップS01の前処理では、この文に対する意味解析木も作成される。

【0033】この単語標準形の表現パターンに対して、ステップS02のマッチングパターンチェックが行われ、この場合は図4のパターンアクションユニット1中のパターン記述部に記述された表現パターンとのマッチングが成立したことがステップS03で判定される。この結果、ステップS04の動作適用処理により、タグ付きテキストの一部

<HD1:Why don't you try new products !>

は、

<HD1:<pref : Why don't you /タグ#1>try new products !>

のように変換される。

【0034】図9は、この動作適用処理後の表示画面の例である。“Why don't you”的部分aは、動作適用処理により“pref”タグでマークされているため、他の部分と異なった様（例えは赤色文字）で表示される。また、この“Why don't you”的部分aは、上記“pref”タグ内の“タグ#1”という情報と関連付けられている。従って、ユーザがユーザデータ入力部104により“Why don't you”的部分を実行指示すると、図4の左下に示す“タグ#1”に基づいて出力処理が行われる。

【0035】図10は、この“タグ#1”に基づく出力処理時のデータ出力部103での表示画面の一例を示した図である。図中“Basic English I”は、図2のパッケージデータ記述部に記述されていたパッケージタイトルである。次の“Why don't you try new products !”は、入力文書中から取り出された部分であり、分かり易くするために、パターン記述部に記述された表現パターンと同一の表現パターンである“Why don't you”的部分が強調表示される。この表示の下に、“Why don't you”的訳文である「～しませんか」という表示がなされ、さらにその下に予め記録されている“Why don't you”的用例が訳文と共に表示される。

【0036】図10の出力処理時の表示画面の最下部には、5つのボタンが示される。「実例」ボタンの部分

は、以前にマッチした実例（この場合、Why don't you を含む例文）へのリンクが張られており、これを押すことにより図4の「実用例」の部分に記録された用例が表示される（図示しない）。「Lesson」ボタンは、「Whyd on't you」という表現の練習を行うための音声ビデオデータやプログラムと関連付けられ、ユーザはこの「Lesson」ボタンを押すことで、会話の練習を行うことができる。「類似」ボタンは、類似表現の関連情報、「反対」ボタンは、反対表現の関連情報を提示する機能をそれぞれ有している。「X」ボタンは、表示画面を閉じる処理を行うためのものである。

【0037】ステップS01の前処理は、図7の<ITM:HUL acquired Ratos Yesterday. >のコンテンツの部分に対しても適用され、単語標準形および意味解析木が内部的に生成される。図11は、この前処理により生成される意味解析木の例を示している。意味解析木は、入力文書に対する辞書引きなどが行われるため、単語に対する各種情報が付与されている。図11では、「HUL」に対して「*ComputerCompany」、「Ratos」に対して「*ComputerSoftwareCompany」という意味カテゴリがそれぞれ付与されている。図4のパタン記述部に記述された表現パタンに対しては、前述のようにマッチングが成功する。この際、「Ratos」に対してパタン記述部に記述された表現パタン中の「*Company」という意味カテゴリがマッチするが、これは「Ratos」に付与されている「*ComputerSoftwareCompany」という意味カテゴリが「*Company」という意味カテゴリとマッチするということによる。パタンアクションユニット2中のアクション記述部には、プログラム#1へのリンクが張られており、この概要が図4のプログラム#1の部分に記述されている。

【0038】図12は、プログラム#1が起動された時の出力処理結果である表示画面の一例である。図12中のaは図9に示した画面であり、bはプログラム#1の提示する画面である。

【0039】図13は、プログラム#1などのプログラムがアクセスする情報源としてのデータベースであり、図12のbの情報は、このデータベースからプログラム#1により取り出され表示されている。図13中のPA1の部分には、プログラム#1により入力文書より抽出されたデータが記憶されている。図13ではPA1の部分に1件のデータのみ記述してあるが、複数件のデータを記憶することも可能である。PA5の部分は、図示しないプログラム#5により記憶されたものであり、これはPA1とは異なったアクションであるが、プログラム#1とプログラム#5は、「コンピュータ会社の会社買収」という共通の抽出対象に入れられており、図12のbの「コンピュータ会社の会社買収」の部分には、PA1とPA5の両方のデータが統合されて表示される。プログラム#5は、例えば「HUL'sacquisition of Rato

s」といった名詞句表現からデータを取り出すというようにし、多くの表現パタンに対応できるようにしたりすることに有効である。

【0040】図12のb中の登録ボタンを押すと、現在、入力文章中から検出されている情報、すなわちHUL, acquire, Ratos, 1995.08 という情報が上記のデータベースに追加保存される。このデータベースは、コンピュータ上のファイルであっても良いし、取り外し可能な記憶媒体に記憶されていても良い。

10 【0041】このように本実施形態では、コンピュータネットワーク上の文書をブラウズしているユーザが、文書中に現れる所望の表現パタンに対して特定の出力処理、例えば、ユーザが見ているページにある英語表現に対する学習データの表示や、その学習データの記憶、といった出力処理を行うことが可能であり、ユーザが欲する情報をしながら言語学習や情報収集を行うことができる。

20 【0042】（第2の実施形態）次に、本発明に係る文書処理システムの一実施形態として、文書処理システムを文書ブラウズ支援システムに適用した実施形態について説明する。

【0043】図14は、本実施形態に係る文書処理システムが要素あるいは付加コンポーネントとして有しているデータ検索システムの構成を示す図である。この検索システムは、文字列を受け取る機能を持つ文字列入力部301と、文字列とコンピュータネットワーク上の文書の位置を示す文書位置情報を保持する文書位置辞書部202と、文書位置情報を出力する出力部303およびこれらを制御する制御部204からなる。

30 【0044】図15は、文書位置辞書部202の内容の一例を示した図である。この文書位置辞書部202は、インターネット上のWWW文書を検索することにより得られる言語表現（文字列）からなる「見出し」と、その言語表現が使用されているWWW文書を特定するためにその文書の位置を示す「文書位置情報」を対応付けて格納したテーブルとなっている。このような文書位置辞書部202を構成する情報は、WWW文書を収集するロボットなどによって集めることができる。図15において、見出しは文字列であり、単語、複合語、句、その他の任意の文字列が入る。本実施形態では、WWW文書にある参照文字列の部分を取り出し、これを見出しとしている。例えば、HTMLにおけるハイパーリンク形式Dyna Book

40 からは、「Dyna Book」が見出しとして取り出され、「http://www.tasuba.com/product/dynabook.html」の部分がそのWWW文書の位置を示す文書位置情報として取り出される。

【0045】さらに、図15の文書位置辞書部202には、「参照」、「更新時」、「更新頻度」の情報も対応

11

付けられている。「参照」は、そのWWW文書に張られているリンクの数である。「更新時」は、そのWWW文書が最後に更新された時間であり、この例では年月日である。「更新頻度」は、例えばここ3カ月における週当たりの更新頻度である。これらの情報は、ロボットなどと呼ばれる既存技術を使用することにより、全世界あるいは、ある特定のネットワーク上の領域のWWWページ全体に渡って集めることができる。これにより、本実施形態では図15の文書位置辞書部202を構成する。

【0046】制御部304は、図16のフローチャートに示す処理を行う。すなわち、まず文字列入力部301によって入力された文字列を取り出し(ステップS1-1)、その入力文字列に関して文書位置辞書部302を検索する(ステップS1-2)。この検索の結果、文字位置辞書部302内に入力文字列にマッチする見出しが存在すれば、その辞書内容である「文書位置情報」、「参照」、「更新時」、「更新頻度」などの情報を出力部303より出力し、見出しが存在しない場合には、その旨を示す情報を出力部303より出力する(ステップS1-3)。従って、文字列入力部301によって文字列を入力することによって、その文字列によりハイパーアリンクを作成している文書がコンピュータネットワーク上に存在すれば、そのハイパーアリンクの先の文書を得ることが可能となる。

【0047】本実施形態の文書処理システムは、このような機能を利用してコンピュータ上の本来のオリジナル文書にはハイパーアリンクが存在しない場合でも、その文書へのリンク情報を新たに提供することができる。以下、この実現法について説明する。

【0048】第1の実現法は、基本的に第1の実施形態で説明した方法と同様の手法を用いる。すなわち、第1の実施形態では、図4のパターンアクションユニット1を用いて、パターン記述部に記述された表現パターンとマッチした入力文書中の表現パターン“Why don't you”に対して、その表現パターンに対応する関連情報としてのハイパーアクション情報を対応付け、図9のaに示したように表示形態上他と異なった表示を行い、この“Why don't you”の部分をユーザデータ入力部104で指示することにより、図10に示したような出力情報をデータ出力部103で表示することでユーザに提示していた。

【0049】この手法を利用して、図15に示すように、文書位置辞書部202のデータ中の見出しに存在する語、例えば“Tasuba”という文字列(第1の実施形態におけるパターン記述部に記述された表現パターンに相当)に、関連情報として文書位置情報であるハイパーアリンク、例えば、

<http://www.tasuba.co.jp/welcome.html>

<http://www.tasuba.co.jp/>

<http://www.abc.com/information/ts.html> 2 4/12/9

5 . 0.003

12

などを対応付けることが可能となる。なお、このようにした場合には、複数のハイパーアリンクをユーザに提示し、これらの中からユーザに選択してもらう必要が出てくる場合がある。

【0050】図17は、ユーザに上記の関連情報を提示する表示画面の例である。図中の“3/3”はWWW文書の数と、それらのうちの表示されている文書に関する数である。図17では図15で示した情報を全て提示しているが、「参照」、「更新時」、「更新頻度」などをパラメータとして、表示する数を制限したり、表示する順番を変更したりすることが可能である。図17の場合、「参照」の多い順に並べている。また、特定の値に表示する数を制限したり、スクロールバーで下位のものも参照できるようにすることもできる。ここでは図示しないが、WWW文書の一部(例えば、表題の文字列)を取り出しておき、それを表示することも有用である。

【0051】また、オリジナル文書において、既に“Tasuba”に対してハイパーアリンクが張られていた場合には、これを上記のようにして新たに対応付けられたハイパーアリンクと合成して表示する。オリジナル文書に張られていたリンク(これをオリジナルリンクと呼び、これと区別するためオリジナル文書に存在しなかったハイパーアリンクを2次リンクと呼ぶこととする)は、重要度が高いので、表示の先頭にする。

【0052】図18は、<http://www.abc.com/information/ts.html>がオリジナルリンクであった場合の表示の仕方を示した図である。図では、オリジナルリンクに“*”を付けて他の異なった様子で表示することにより、オリジナルリンクであることを表している。オリジナルリンクは大事であるため、常に表示するという様子をとってもよい。

【0053】このようにして、ユーザはコンピュータネットワーク上のWWW文書を取り込んだ際に、オリジナル文書には付いていなかったハイパーアリンクをも容易に検索して表示し、それを利用して文書ブラウジングを行うことが可能となる。

【0054】本実施形態では、文書ブラウザで取り込んだ文書の内容を書き替えることにより、オリジナル文書にないハイパーアクション情報を検索・表示・利用可能としている。しかし、この方式だけが実現法というわけではない。例えば、オリジナル文書の内容を変更することなしに、同様の効果を取り出すことが可能である。これには、例えばブラウズ表示を行うコンポーネントにおいて、オリジナル文書の情報とハイパー情報込みの文書の情報の2つないしはそれ以上を管理すれば良い。

【0055】(第3の実施形態) 次に、本発明に係る文書処理システムの他の実施形態として、文書処理システムを文書作成編集システムに適用した実施形態について述べる。コンピュータネットワーク向けの文書を作成する場合は通常、ネットワーク上の他の文書に対してハイ

13

パーリンクを張ることを行う。一般に、この作業にはリンクを張りたい文書のネットワーク上の位置を示す文書位置情報を入力することが必要であるが、この際に行うネットワーク上の文書の検索には、従来、かなりの熟練と時間を要していた。

【0056】本実施形態では、ハイパーリンク付き文書の文書作成システムにおいて、第2の実施形態で説明した2次リンクを一時的に付与する機能を用意し、文書作成者の指示した2次リンクのみをオリジナルリンクとする機能を用意する。そして、作成している文書中にハイパーリンクを付与したい文字列、例えば“Tasuba”という単語が存在したとする。マウスによるドラッグなど既存の文字列指定技術を利用して、あるいは設定可能なハイパーリンクを持つ語を区別して表示して、その文字列を指示することにより、ユーザは文字列を指定することができる。

【0057】この文字列に対する設定可能なハイパーリンクは、第2の実施形態と同様に、図17に示したような形態でユーザに指示することが可能である。これらのハイパーリンクの候補から、マウスなどの入力手段を利用して、設定したいハイパーリンクを適当に指定することで、設定すべきハイパーリンク情報を選択できる。このハイパーリンク情報をマークアップ言語などの所定のハイパーリンクのデータ形式で設定することは、公知技術によって実現が可能であり、これによりユーザは、自分で作成した文書に他の文書で設定されているハイパーリンクを参照して、ハイパーリンクの選択・付与を行うことができる。

【0058】なお、本発明は上記の実施形態に限定されるものではなく、日本語やフランス語など、任意の言語に対して適用することが可能である。要するに、本発明はその主旨を逸脱しない範囲で種々変形して実施することができる。

【0059】

【発明の効果】以上説明したように、本発明によれば少なくとも文書中の表現パターンとそれに対応する出力情報を得るための該出力情報に関する関連情報を記述したデータを入力し、このデータ中の表現パターンを参照してコンピュータネットワーク上の文書中の表現パターンを検出して、該データ中の関連情報を基に、検出した表現パターンに対応する出力情報をコンピュータネットワーク上の文書から生成して出力することによって、コンピュータネットワーク上の文書をブラウジングする際、文書作成者以外の人によりブラウジング対象の文書への関連情報の付与を行うことが可能であり、コンピュータネットワーク上の文書をブラウズしているユーザが文書中に現れる所望の表現パターンに対して特定の出力を行うことができる。具体的には、例えば、見ているページにある英語の表現に対する学習データの提示や、その学習データの記憶といった出力処理を行うことができ、ユーザが自

14

分の好みの情報を見ながら学習や情報収集を行うことができる。

【0060】また、少なくとも文書中の表現パターンとコンピュータネットワーク上の該表現パターンを含む文書の位置を示す文書位置情報を対応付けて記述したデータを入力し、このデータ中の表現パターンを参照してコンピュータネットワーク上の文書中の表現パターンを検出して、この検出した表現パターンに対応する文書位置情報を入力されたデータから検索して表示などの出力をすることにより、コンピュータネットワーク上のブラウジング対象の文書中に文書位置情報であるハイパーリンクが存在しない場合でも、ハイパーリンクを対応付けて出力することが可能となる。

【0061】さらに、文字列とコンピュータネットワーク上の該文字列を含む文書の位置を示す文書位置情報を対応付けて格納しておき、入力された文字列を基に文書位置情報を検索して表示などの出力をすることにより、ハイパーリンクを含む文書を作成する場合、文書位置情報として、他者が文書中に作成したハイパーリンク情報を検索し表示して利用することが可能となる。

【図面の簡単な説明】

【図1】本発明の一実施形態に係る文書ブラウズ支援システムの概略構成を示すブロック図

【図2】同実施形態で利用するプラグインパッケージのデータの概要を示す図

【図3】図2中のパターンアクションユニットを参照実行するアルゴリズムを示すフローチャート

【図4】図2中の2つのアクションユニットの具体例を示す図

【図5】同実施形態における他のネットワーク上の文書へのハイパーリンクを含む文書をブラウザに表示する時の表示画面の一例を示す図

【図6】同実施形態におけるブラウズ画面の一例を示す図

【図7】同実施形態における送出された文書データの一部を示す図

【図8】同実施形態における単語標準形に対して得られる情報を示す図

【図9】同実施形態における表示画面の一例を示す図

【図10】同実施形態における表示画面の一例を示す図

【図11】同実施形態における意味解析木の一例を示す図

【図12】同実施形態における表示画面の一例を示す図

【図13】同実施形態における記憶情報の一例を示す図

【図14】本発明の他の実施形態に係る文書処理システムの概略構成を示すブロック図

【図15】同実施形態における文書位置情報の一例を示す図

【図16】同実施形態における制御部の動作を示す図

【図17】同実施形態における情報提示例を示す図

10

20

30

40

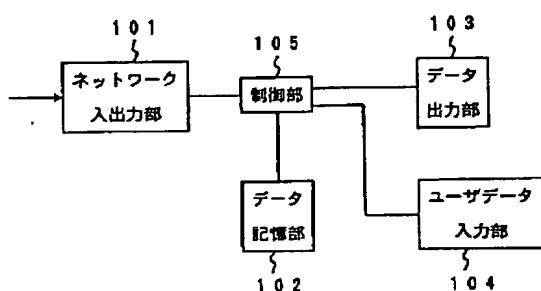
50

【図18】同実施形態における記憶情報の一例を示す図

【符号の説明】

101…ネットワーク入出力部
102…データ記憶部
103…データ出力部
104…ユーザデータ入力部

【図1】



【図3】

105…制御部

201, 202, 203…パタンアクションユニット

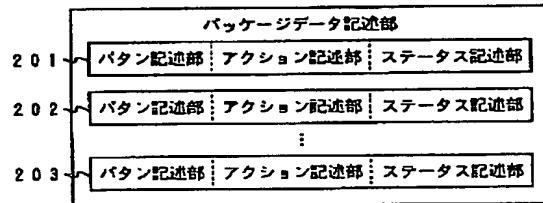
301…文字列入力部

302…文書位置辞書部

303…出力部

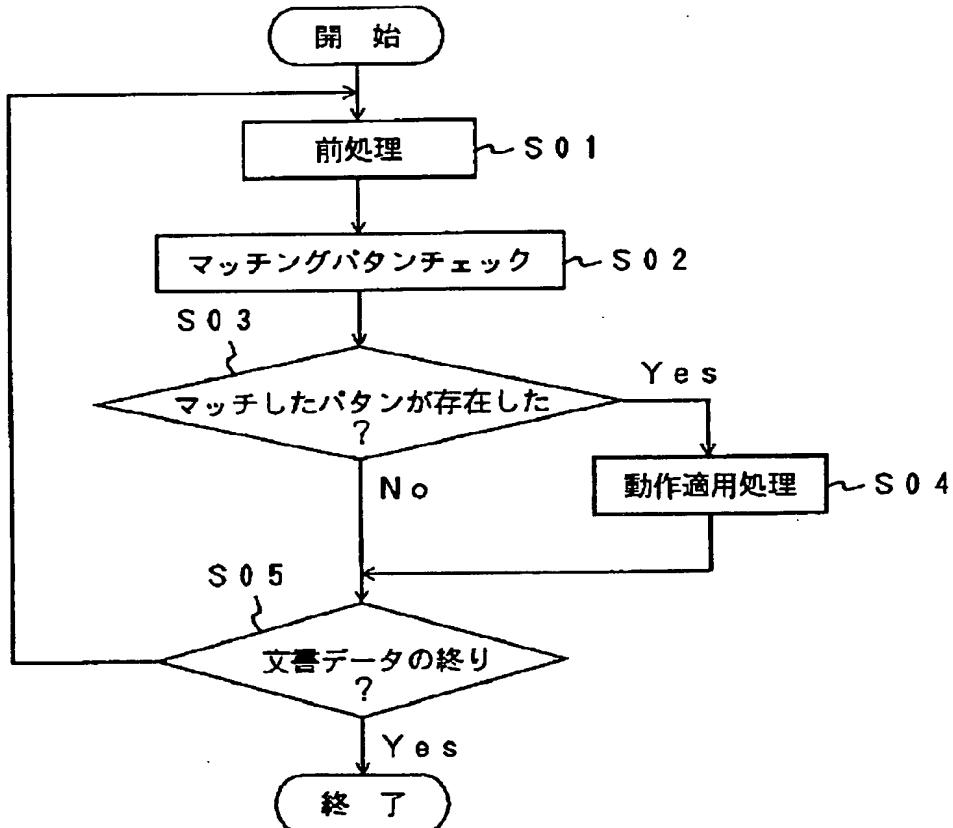
304…制御部

【図2】

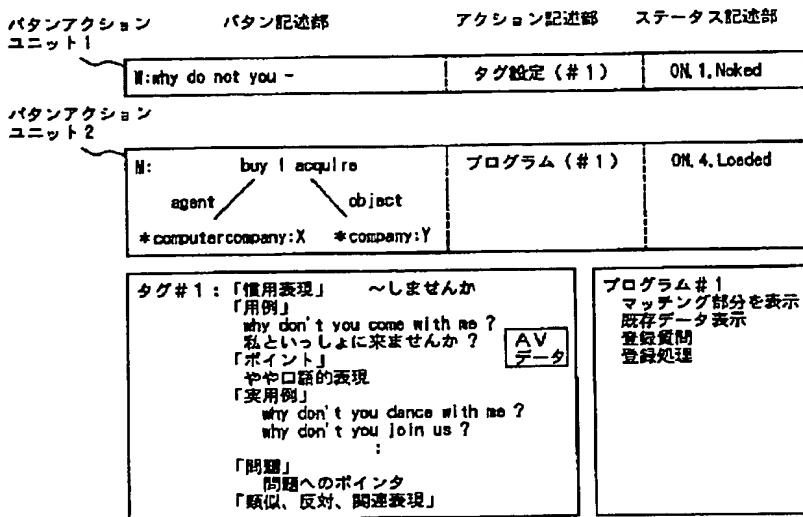


【図8】

1 why (NOUN, HeadCap) 45-47
2 do (AUX) 49-
3 not (NOT) *-53
4 you (PRO) 55-57
5 try (VT) 59-61
6 new (ADJ) 63-65
7 product (NOUN, p!) 67-73



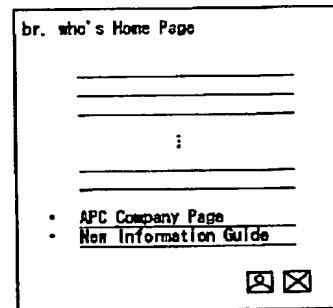
【図4】



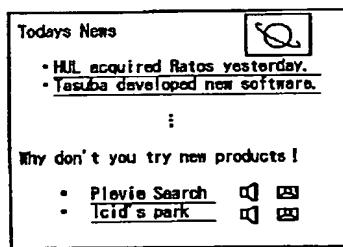
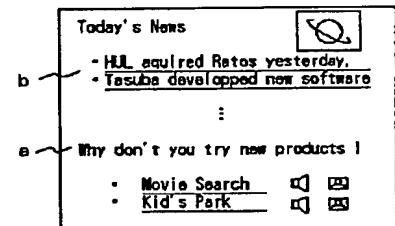
【図6】

【図7】

【図5】



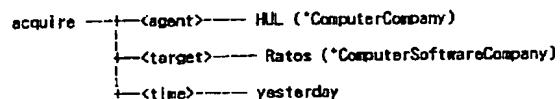
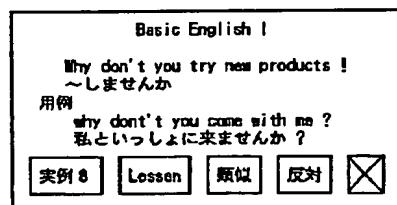
【図9】



<TEXT:
 <HD1: Today's News>
 <ITM1
 <ITM: HUL acquired Ratios yesterday.>
 <ITM: Tesuba developed new software.>
 :
 >
 <HD1: Why don't you try new products !>
 <ITM1
 <ITM: Movie Search><OBJ: speaker1><OBJ: person1>
 <ITM: Kid's park><OBJ: speaker2><OBJ: person2>
 >

【図10】

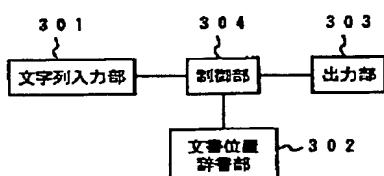
【図11】



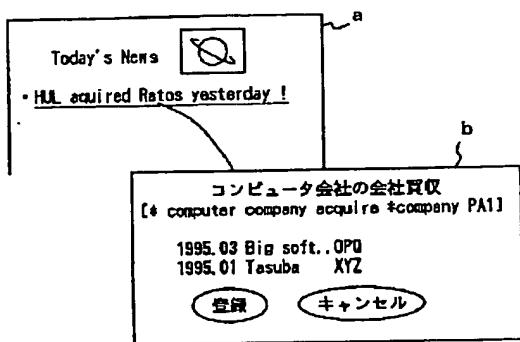
【図13】

コンピュータ会社の会社買収				トムホーキンス		
PA1	Bigsoft	acquire	Intra	1995.03	PA23	buy visit new house Japan China
PA5	Tesuba	buy	XYZ	1995.01		

【図14】



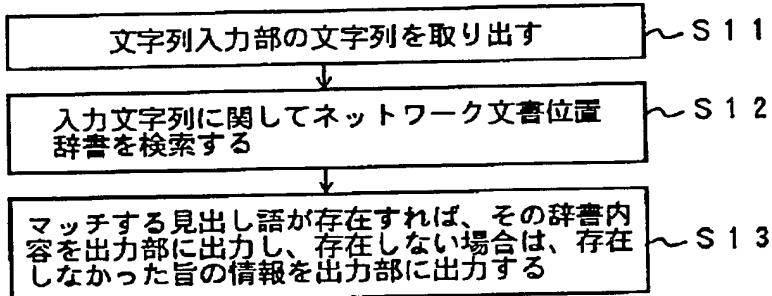
【図12】



【図15】

見出し	位 置	登録	更新時	更新頻度
ABC :	http://www.abc.com/ http://www.abc.com/intro	456	6/20/96	0.6
	http://www.toycompany.com/prd/abc.html	30	4/20/96	0.1
ABCD :	http://www.abcd.com/	2	6/13/95	0.001
		23	3/23/96	0.8
Toshiba :				
	http://www.toshiba.co.jp/welcome.html	4000	5/28/96	2.2
	http://www.toshiba.co.jp/	20	5/23/96	1.2
	http://www.abc.com/information/ts.html	2	4/12/95	0.003

【図16】



【図17】

3/3 Toshiba	10
4000 5/28/96 2.2 http://www.toshiba.co.jp/welcome.html	
20 5/23/96 1.2 http://www.toshiba.co.jp/	
2 4/12/95 0.003 http://www.abc.com/information/ts.html	

【図18】

3/3 Toshiba	10
2 4/12/95 0.003 http://www.abc.com/information/ts.html	
4000 5/28/96 2.2 http://www.toshiba.co.jp/welcome.html	
20 5/23/96 1.2 http://www.toshiba.co.jp/	

フロントページの続き

(51) Int. Cl. 6

識別記号

F I

G 0 6 F 15/62

3 3 0 G